

A Review of the Document: Development of a Site-Specific Iron Water Quality Criterion for the Chuitna River Drainage, Alaska.

National Center for Environmental Assessment, ORD, EPA

Summary

In general, the analysis and assessment performed by TetraTech was well done given the limitations of the data available to them. However, we believe that there is still substantial uncertainty concerning the background concentrations of iron and the sensitivity of the biota.

We believe it is advisable to require a well designed program of a biological and chemical (iron and other water quality parameters including the full ionic matrix) monitoring as part of any decisions to allow a variance in the iron standard or grant a mining permit. This could be part of an adaptive management process. Specific technical comments follow.

Estimation of background

Recognizing the importance of site-specific conditions, EPA allows for setting site specific water quality criteria if there is a site-specific reason. That is, if it can be credibly shown that the undisturbed background level of the regulated agent is naturally greater than the default criterion or that the local water characteristics or the species composition of the receiving community affect the toxicity of the criterion chemical. However, it must be credibly shown that, when using the proposed site-specific criterion, species receive the same or greater level of protection as provided by the existing criterion.

To demonstrate that the undisturbed background level of the regulated agent is naturally greater than the State standard, the applicant must show that iron occurs naturally in greater concentrations. The paucity of data provided for the watershed is not sufficient temporally or spatially to establish natural background levels for iron. This is demonstrated by the fact that the authors resorted to averaging across the watershed. Some measurements show that portions of the basin have low iron levels. Water quality studies should identify tributaries and seasons in which iron is elevated and those in which it is low.

Because the variance among individual tributaries is not characterized, the effects of raising the standard on iron levels in the river are not clear. If naturally low iron tributaries become high iron tributaries, the assimilative capacity may not be sufficient to protect downstream reaches if the dilution provided by low-iron tributaries was lost.

Where iron levels were greater, the authors did not convincingly demonstrate that the sources were natural. The report's statement on this issue is: "A limited amount of oil and gas exploration and logging has occurred in the basin, but reportedly not in a quantity that is expected to impact the natural conditions of surface water." The wording of this statement suggests that the authors did not have direct access to information and the source is not reported. We suggest that, if good data are not available concerning activities in the watershed, a survey be performed for potential sources, focusing on tributaries with high iron levels.

Exposure events that may be toxic are masked by averaging concentrations. Undoubtedly, this was not the intent of the contractor, but the concentrations were averaged because the dataset was not optimal for a proper analysis. Particularly, the biological data was not paired with water quality data. This underscores the inadequacy of the data set for the development of site-specific criterion development. This inadequacy of the data cannot be overcome by statistical manipulation.

Finally, the chemical form of natural iron and of any iron-bearing effluents may not be the same. As the authors acknowledge, the aqueous chemistry of iron is complex. If an effluent has different chemistry, particularly lower pH, the iron concentration may be within the background range, but the background may not be relevant.

Estimation of Effects

Even if background levels are high, site specific criteria should not be harmful. Even if toxic conditions occurred naturally in a few locations, one would not revise a criterion to allow toxic conditions everywhere. Hence, the analysis of effects is important even if the background range was shown to include high iron concentrations.

Water quality criteria are intended to be protective of all but the most sensitive species with respect to the contaminant of concern. The aggregate biological assessment metrics that were presented are intended to respond to most common contaminants and may be poor measures of responses to ions. The mathematical result of lumping species together hides which species are being lost and what is replacing them. For example, the authors state: "If iron concentration was impairing macroinvertebrate community in the Chuitna basin, EPT taxa should be lower with higher iron concentration." However, it is well known that some EPT taxa are tolerant of metals and will increase in the presence of some effluents. It is akin to comparing one backyard with a rabbit, deer and a squirrel and another with a rabbit, a squirrel, and 20 rats. Both have three species in their backyard, but they are qualitatively different. In sum, the metrics used in the report are useful for showing impairment, but not for setting a level that would protect 95% of species (Stephen et al 1985) or that would protect individual highly valued salmonid species or functionally important invertebrate species.

Since the national iron criterion was not developed using toxicity test measurements, the applicant is reasonable in attempting to define effect levels using field data. However, the data set used is inadequate to the task. A larger data set from paired biological and water quality samples is required to confidently refute or discount the iron standard.

As the authors explain, not all iron moieties are similarly toxic. This makes the task of developing a site-specific iron criterion more difficult. As discussed above, the form of iron in future effluents and their receiving waters may differ qualitatively from the background iron. In addition, the physical effects of fine particulate iron described by the authors do not have well defined kinetics and dynamics. It may be that those effects occur in relatively short episodic exposures or relatively long average exposures. That issue affects the averaging times and sampling frequencies required to characterize relevant iron exposures.

The guidance for developing site specific criteria from field data is less well established than the toxicity test derived methodology. Setting the criterion at the background level may be appropriate, but the appropriate expression of background needs consideration. The authors calculate means and maxima and use maxima for the criterion but it is not

clear what data set is averaged or searched for a maximum value. Within the report, the set definition seems to vary from sites, to streams, years and the entire watershed. This must be clarified. Certainly, the maximum observed value across years and sites in the watershed is not sufficiently protective. It would be inappropriate to use that maximum iron value unless it was paired with the poorest biological value and there was no effect. The point is, exposure and effects should be matched. Unless the authors can demonstrate that there were not biological effects from the maximum iron event at the time and location where it occurred, then they have not justified its use.

The authors' use and interpretation of hypothesis testing statistics is inappropriate and is based on a common but fundamental error (Sec. 2.3). You cannot conclude that no difference exists based on failing to reject the null hypothesis. Null hypothesis testing requires that you assume the null hypothesis and then determine the probability of data as extreme as those obtained or more extreme, given that the null hypothesis is true. Clearly, failing to reject is not the same as accept. Also, hypothesis tests depend as much on the number of replicates as on the differences between treatments. This was not a problem for Fisher, because he assumed that you could always gather more data if the results were ambiguous. However, in this case there are few data and the authors did not have the option of collecting more. Finally, there are no replicated treatments and no random assignments of treatments, so the observations are pseudoreplicates.

The use of exposure-response modeling for fish is more appropriate (putting aside the p statistic), and it does appear that there is no relationship. However, 7 is very few points, and aggregate metrics are used. The use of % salmonids alone in a coldwater system where the natural assemblage is 88-99% salmonids is inappropriate. Even in places where a taxon of interest is not so dominant, % metrics are problematical. A decline in % salmonids could occur, because conditions improved for a non-salmonid species. It would be better to examine the responses of individual species, particularly since there are relatively few in this system. Also, the independent and dependent variables are reversed in Fig. 7. Depending on the regression method used, this can make a difference in the result.

A method for deriving criteria from field data, which was developed with the assistance of TetraTech, was recently reviewed and commended in draft comments by a panel of the EPA Science Advisory Board (U.S. EPA 2010). (Note, this method has not completed final review by the Board, but is expected in early 2011). The method could not be applied to the existing data and may not be applicable to iron at this site even with more data. However, it illustrates that field-based methods can address the responses of individual sensitive taxa and therefore can be equivalent to the standard criteria derivation method.

Implications

It seems to us that there are at least three options for dealing with this situation.

1. Accept the proposal to develop a site specific criterion using the existing data. This should include some revision and clarification of the assessment. In particular, the derivation of the maximum values and their relations to specific associated biological effects should be explained. Given the small data sets (28 for invertebrates and 7 for fish) and the rough spatial association between chemical and biological data, there is considerable uncertainty associated with this option.

2. Agree that a site-specific criterion may be appropriate but demand better data and analyses. The sample set should contain numerous distinct sites distributed in the basin and adjoining basins. Biological and chemical samples should be co-located in space and time and should determine the variance among tributaries. Sampling should bracket seasons of emergence of sensitive invertebrate taxa. The low human population density in the region should provide sampling sites from which very credible measures of background levels of iron and other water quality parameters can be estimated.

3. Address the issue of iron and all other constituents of mine effluents by using an adaptive management approach. Chemical and biological monitoring prior to and during mining could determine whether iron is elevated relative to site-specific background and could detect biological effects. Problematic changes in water quality and biology could lead to improvements in management of aqueous effluents.

Specific Comments

The authors state: p. 4.

“The goal of these analyses is to identify the highest iron concentration at which the biological community is not impacted similar to a toxicological No Observed Effect Concentration (NOEC) which EPA uses in development of other toxics criteria (Stephen et al 1985). If no relationship is observed between iron concentration and biological community response, it may be concluded that observed iron concentrations do not have an impact on the biota in this system.”

Page 7.

“The maximum 4 mile distance between iron and macroinvertebrate sites ensures that macroinvertebrate data are related to iron data used in analyses.”

The mayfly number or % is the most sensitive metric provided and it shows an effect, but it is confounded and due to the small sample size cannot be deconstructed. The signal from the popular metric, EPT can be swamped by tolerant Tricoptera. It would be good to provide the taxa list for verification of the abundance of *Hydropsychidae* and *Cheumatopsyche*. As noted previously, the numbers of samples are insufficient for confidently establishing a background or effects assessment (Page 5, “2006 to 2008 (12 vs. 28 stations)”). Without paired biological and chemical data, associations are not convincing.

The authors state:

“In these comparisons, the July 2007 macroinvertebrate results were not significantly different from the July 2006 results even though average and maximum total iron concentrations increased from 1.5 and 2.0 mg/L, respectively in July 2006 to 2.0 and 3.0 mg/L in July 2007. (Figure 6 and Appendix D). These results indicate that observed iron concentrations are unrelated to macroinvertebrate community integrity”

This strongly stated conclusion from only two sampling dates ignores the possibility that the invertebrates may be responding to peculiar events in a particular year and may be responding to conditions in the previous year because the early life stages are more sensitive than the larger organisms that are collected.

The authors state: “In fact, all biological data examined in these analyses indicate that the biological integrity in this basin is high and similar to biological reference conditions identified by ADEC et al. (2007) for the Cook Inlet region.” However, they present results only for % Ephemeroptera.

In the following sentence: “A low R^2 and high p value (e.g., > 0.10 or 10%) would indicate no relationship between iron concentration and fish community integrity.” The expression in parentheses does not make sense.

High variance suggests that the data should be corrected for confounders such as stream size and seasonality.

References

U.S. EPA. 2010. A Field-based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams, External Review Draft. National Center for Environmental Assessment, Washington, DC; EPA/600/R-10/023A. accessed on 11/21/10 at http://oaspub.epa.gov/eims/eimscomm.getfile?p_download_id=495377